Cell
Leading Edge
Forum

# Metabolic Phenotyping in Health and Disease

Elaine Holmes,[1] Ian D. Wilson,[2] and Jeremy K. Nicholson[1,*]
[1]Department of Biomolecular Medicine, Division of Surgery, Oncology, Reproductive Biology and Anaesthetics, Faculty of Medicine, Imperial College London, South Kensington, London SW7 2AZ, UK
[2]Department of Safety of Medicines, AstraZeneca Pharmaceuticals, Mereside, Macclesfield, Cheshire SK10 4TG, UK
*Correspondence: j.nicholson@imperial.ac.uk
DOI 10.1016/j.cell.2008.08.026

**Analyzing metabolites (small molecules <1 kDa) in body fluids such as urine and plasma using various spectroscopic methods provides information on the metabotype (metabolic phenotype) of individuals or populations, information that can be applied to personalized medicine or public healthcare.**

A major challenge for 21st century medicine is to understand the relationships and interactions between genetic variations and environmental triggers of disease (Nicholson, 2006). In response to this challenge, systems biology powered by genomics, proteomics, bioinformatics, and now metabolomics and metabonomics (Box 1), is providing a new logical framework to elucidate disease etiology and to uncover latent connections between seemingly disparate disease states (Loscalzo et al., 2007). In particular, genome-wide association studies (GWAS) enable genetic features to be linked to many pre-disease conditions (Kruglyak, 2008). But GWAS can explain only a small proportion of incidence variation and may not translate between different populations—for example, having two copies of an *FTO* gene variant is associated with significantly higher levels of obesity and type 2 diabetes in Western populations but not in the Han Chinese (Li et al., 2008a). Furthermore, the relationships between individual genomic and phenotypic variations in response to drug treatment is still poorly understood, and it is unlikely that genetic information alone will be able to direct personalized drug therapy.

The metabotype or metabolic phenotype (Box 1) provides a readout of the metabolic state of an individual and is the product of genetic and environmental (diet, lifestyle, gut microbial activity) contributions under a particular set of conditions (Nicholson, 2006). Analyzing metabolites (small molecules <1 kDa) in body fluids such as urine and plasma using various spectroscopic methods provides knowledge of the metabotype. Such metabolic profiles provide information that cannot be obtained directly from the genotype, gene expression profiles, or even the proteome of an individual. These metabolic profiles provide a top-down "systems level" readout of the biochemistry, physiological status, and environmental exposure of individuals and populations that can be exploited in personalized medicine and public healthcare (Nicholson, 2006). Metabolism-driven approaches should prove highly tractable as they combine the gathering of systemic information based on minimally invasive analysis with high-throughput capabilities.

## Metabolome Variation and Metabolic Profiling

Factors such as gender, age, diet, gut microbiota, physical activity, latent disease, medication, hormone, and stress levels modify the metabotypes of both individuals and populations (Figure 1), and hence the prevalence and risk of disease (Nicholson, 2006; Daviglus et al., 2004; Li et al., 2008b). Metabolome-wide association studies (MWAS) have the capacity to link human metabotype variations to disease risk factors in the general population (Holmes et al., 2008). Xenobiotics such as plasticizers, food preservatives, pesticides, and plant secondary metabolites (caffeine, flavanoids, phytoestrogens) also contribute to metabotype variations as does gut

**Box 1. Glossary**

**Conditional metabolic phenotype:** A characteristic metabolite profile reflecting the host genome and its interaction with environmental factors, diet, and the gut microbiome.
**Co-metabolite and co-metabolome:** Compound or set of compounds derived from interactions of more than one genome in symbiotic systems.
**Metabolome:** A quantitative description of all endogenous low-molecular-weight components (<1 kDa) in a biological sample such as urine or plasma. Each cell type and biological fluid has a characteristic set of metabolites that reflects the organism under a particular set of environmental conditions and that fluctuates according to physiological demands. The metabolome can be divided into the primary metabolome (controlled by the host genome) and the co-metabolome (dependent on the microbiome).
**Metabonome:** Theoretical combinations, sums, and products of the interactions of multiple metabolomes (genome, symbiotic, parasitic, environmental, and co-metabolic) in a complex system.
**Metabonomics:** The quantitative measurement over time of the metabolic responses of an individual or population to drug treatment or other intervention.
**Microbiome:** The consortium of microorganisms, bacteria, protozoa, and fungi that live commensally or symbiotically with a host.
**Pharmacometabonomics:** Prediction of the quantitative outcome of a drug intervention in an individual based on a pre-dose mathematical model of their metabolic state.
**Xenometabolome:** Characteristic profile of nonendogenous compounds (drugs and their metabolites, pollutants, dietary components, herbal medicines) in a biofluid.

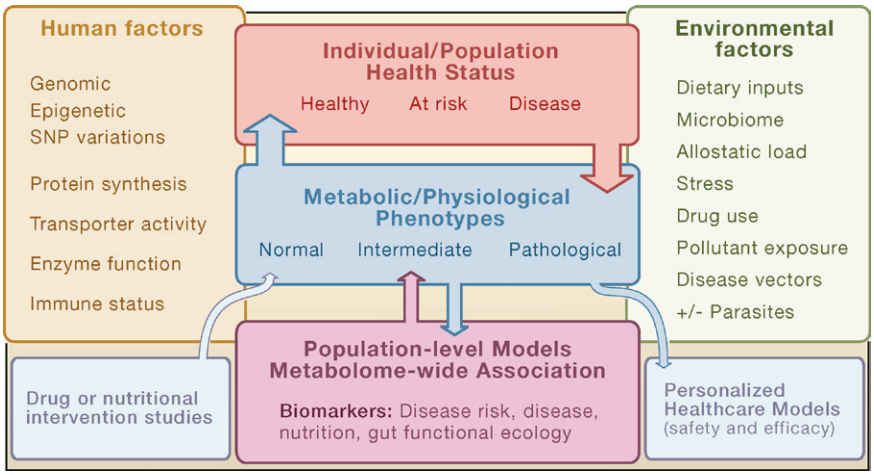


**Figure 1. Interactions between Genes and the Environment and Their Effects on Health**
Metabolic phenotypes (metabotypes) can be measured by profiling small molecules (<1 kDa) in biofluids (usually urine or blood) by spectroscopic techniques such as NMR. Metabolic phenotypes are influenced by intrinsic and environmental factors that determine health status and disease risk of an individual or group. Measuring and modeling the profile of all metabolites (metabolome) in an individual may provide insights into disease risk factors and etiology, information that could be used for personalized medicine.

microbiome activity through the production of co-metabolites (Li et al., 2008b; see Essay by Turnbaugh and Gordon, page 708 of this issue). The composition of the microbiome (Box 1) has an impact on a variety of disorders from intestinal disease to obesity and cancer. Understanding variation of the co-metabolome (Box 1) will shed light on how an imbalance in our microbiota may contribute to disease.

Most analytical platforms for metabolic profiling are based on spectroscopic techniques, e.g., nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry (often combined with chromatographic separation), to permit extensive and rapid analysis of small molecule metabolites, generally with minimal sample preparation (Nicholson et al., 2002). These approaches provide relatively high-throughput analyses at low cost per sample. Multiparameter datasets containing quantitative information on a comprehensive range of classes of small molecules can be produced, often in a nontargeted manner. Given the complexity and high-density nature of the spectral data, many computer-based preprocessing and multivariate modeling techniques have been developed to facilitate the analysis and interpretation of these data, in terms of the response of an organism to physiological or pathological events; such bioinformatics techniques can also be used to probe gene-environment interactions. Mathematical modeling of metabolic data enables diagnostic characterization allowing identification of potential biomarkers, which may contribute to a better understanding of pathological processes and potentially to the elucidation of new drug targets (Nicholson et al., 2002).

Discovery of new metabolic biomarkers using untargeted metabolic profiling could be translated into clinical tools for application to personalized medicine. For example, we have demonstrated proof-of-principle of pharmacometabonomics (Box 1), in which a mathematical model of a group of pre-treatment metabolite profiles in rats was used to predict post-drug interventional outcomes such as toxicity and metabolism of the xenobiotic itself (Clayton et al. 2006). Human variability is greater than that of experimental animals and this poses problems as well as advantages for modeling, but in principle the greater the variation in the metabolic starting point the greater the variation in resulting observed responses. Clearly, in a clinical setting, the complex analytical instrumentation used in the discovery phase of this type of work would not be appropriate or affordable for the measurement of these markers in patients. Implementation of this type of monitoring will depend on evolving technologies such as "lab-on-a-chip" methods, sensors and electronic "noses," which can detect complex molecular fingerprints (Westhof et al. 2007) that can be tuned to detect specific biomarkers (or pathways) rather than providing comprehensive metabolic profiles. Ultimately, the goal is to have handheld "intelligent" devices in the clinic, possibly incorporating microfluidics, that would detect selected proteins, mRNAs and metabolites simultaneously in a pin-prick of blood. An emerging area that illustrates how this might be achieved is in breath analysis using ion mobility spectrometry (IMS), which has shown promise for the diagnosis of infections, lung cancer and sarcoidosis based on detecting a range of low molecular weight volatile metabolites in exhaled breath (Westhof et al., 2007).

Another potential strategy is to measure the chiroptical properties of metabolites, an interesting approach given that different organisms use different forms of optically active isomers. Indeed, symbiotic organisms including man contain both L- and D-forms of amino acids reflecting their human and gut microbial sources. Thus "chiral metabonomics" using chiroptical spectroscopy or chiral chromatographic derivitization methods can be used to differentiate small molecules originating from mammalian and microbial metabolism, potentially helping to elucidate how such symbiotic interactions may go awry in intestinal diseases. For example, L-lactate is a mammalian product of anaerobic metabolism, but gut bacteria produce excessive D-lactate in many gastrointestinal disorders causing systemic acidosis that cannot be detected with the standard lactate dehydrogenase assay, which only detects the L-lactate stereoisomer. Methods designed to quantitate chiral metabolic components will help to deconvolute human and symbiont contributions to the metabotype and their relative importance in the etiology of the condition or response under study.

## Applying a Systems Biology Approach to Medicine

Systems biology is providing many new tools to describe, model, and visualize the integrated action of regulatory networks at all levels of biological organiza-

tion (Loscalzo et al., 2007). In "top-down" systems biology we study how biocompartments (subcellular, cellular, tissue, organ) interact in space and time using minimally invasive methods that capture the properties of systemic homeostasis and its dysregulation (Nicholson, 2006). The efficient integration of "omics" data and platform outputs via bioinformatics should provide a deeper understanding of multilevel connections including gene-gene and gene-environment interactions involved in disease development. The next decade will see the translation of these tools and approaches into predictive and preventive "systems medicine."

By extracting lists from gene expression profiles and metabolic profiles associated with disease or therapeutic intervention, disparate biological events can be connected. Several approaches have been applied to link genotypes and phenotypes in animal models. Examples include reverse genetics, in which a single gene is ablated and the phenotypic consequences ascertained, and forward genetics, which exploits subtle allelic variation relying on identification of mutant strains (and underlying genes) from the phenotype. Of particular interest are methods that investigate gene-gene interactions directly by analyzing and interpreting microarray expression data against a background of traditional genetic mapping techniques to establish associations between quantitative trait loci and expression data (eQTLs). Three susceptibility genes for obesity were identified by an eQTL approach that predicted transcriptional responses to a single gene perturbation in mice (Schadt et al., 2005). This principle was subsequently extended to the use of metabolic profiles (mQTL) as the basis for mapping loci associated with quantitative changes in the plasma of a rat strain susceptible to type 2 diabetes (Dumas et al., 2007). An alternative approach to forward genetics has recently been proposed that involves identification of gene networks that are perturbed by susceptibility loci that may lead to the development of disease. This approach allows the emergent properties of gene relationships to be discovered (at least in experimental animals) and the validation of new genes associated with disease risk such as those associated with obesity (Chen et al., 2008).

Tools for modeling metabolic networks include those in which nodal architectures (corresponding to genes, gene products, proteins, or phenotypes) are connected to demonstrate interactions between those nodes. More sophisticated approaches using hierarchical modularity network strategies reflect the clustering of nodes and interconnection probabilities (Loscalzo et al., 2007). An alternative approach to integrating multiple omics datasets is to coanalyze the disparate matrices to identify those data that co-vary and those that are unique. It is important to elucidate how different levels of cellular nanomachinery interact in space and time to control metabolism and how they are disrupted in disease. A variety of advanced mathematical tools including linear projection or Bayesian (conditional probabilistic) methods with bidirectional orthogonal filters (to remove unwanted noise components in the data) can be used to identify covariation between metabolite profiles and transcriptomic, proteomic, or metagenomic data matrices and so give new insight into multilevel interactions in whole organisms and how they relate to disease (Nicholson, 2006).

## Predicting the Outcome of Therapeutic Interventions

A key goal of modern medicine is to develop personalized therapies tailored to an individual's biology. Patient stratification based on biology (genetic and/or phenotypic) is usually viewed in terms of maximizing drug safety and efficacy. Individuals respond differently to therapeutic interventions, and adverse drug reactions are a cause for concern, especially idiosyncratic toxicity that is revealed when large populations are exposed to new therapeutics. Variations among individuals in response to therapies are influenced by differences in the conditional metabolic phenotype (Box 1). In the mathematical modeling of these interindividual variations, the conditional metabolic phenotype can be thought of as a starting point (pretreatment) for an individual in a conceptual multivariate metabolic space that reflects the combination of many physical, chemical, genetic, and environmental influences. It is this starting position (irrespective of relative contributions of the individual "vector" components) that determines the outcome of an intervention. If the factors that predispose individuals to such idiosyncratic events could be determined, then at-risk individuals could be screened out of a patient population helping to retain pharmaceuticals that are of value to the majority of patients.

Most personalized approaches to drug treatment so far have been based on measuring genotype variation and polymorphisms in drug metabolizing enzymes, such as cytochrome P450 and *N*-acetyl transferases (Nebert et al., 2003). However, pharmacogenomic predictions of drug metabolism and toxicity have proved disappointing. We have demonstrated a pharmacometabonomic approach to predicting the interventional outcome of drugs based on mathematical models of pre-dose metabolic profiles (Clayton et al., 2006). We used a projection-to-latent structure (PLS) modeling method to predict liver toxicity in rats given a threshold toxic dose of the analgesic acetaminophen. There was a significant association between the pre-dose metabolite profile in rat urine and the post-dose outcome regarding urinary excretion of acetaminophen metabolites and the severity of liver damage. Although preliminary, this work shows that there is potential for applying pharmacometabonomics non-invasively for screening human populations. However, a better understanding is needed about the relationships between endogenous metabolic status and drug metabolism outcomes for a wider range of drugs. Ultimately, a judicious mixture of pharmacogenomics and pharmacometabonomics will be required to meet the personalized healthcare challenge of individualized drug therapy.

## Molecular Epidemiology and Metabolome-Wide Association Studies

There is considerable geographical metabotype variation between population groups linked to ethnicity and lifestyle, and this diversity may be closely related to disease risk and incidence (Holmes et al., 2008). There are well-documented disparities in genetic background and diet between geographically dispersed populations, and there are also regional variations in the incidences

and concentrations of foreign (xenobiotic) metabolites. These xenobiotics also contribute to the human metabotype and its regional variation in the form of the foreign compound profile or xenometabolome (Box 1) of an individual or population.

The goal of MWAS is to perform large-scale metabotyping of populations using spectroscopic methods and to relate these metabotypes to disease risk factors. The primary advantage of this approach is that the resulting biomarkers are genuine metabolic end-points and investigations into these pathway perturbations may yield new therapeutic targets or insights into dietary interventions. This has been demonstrated in human populations: a biomarker (formate) identified in the human urinary metabolome inversely correlates with blood pressure (Holmes et al., 2008). Dietary salt intake is strongly linked to blood pressure; formate is involved in chloride-ion exchange in the kidney via the chloride-formate exchanger and the SLC26 transporters (a series of interconnected renal ion exchangers involved in $Na^+$ and $Cl^-$ transport) (Sindic et al., 2007). In this case, there is a clear connection between the MWAS-derived metabolic biomarker and physiological evidence pointing to its role in blood pressure regulation. Thus, MWAS studies have the potential to provide new insights into disease mechanisms and pathophysiology that may ultimately lead to new drug targets. As there are many hundreds of major epidemiological data sets and sample repositories worldwide with known long-term outcomes, there is a great opportunity to use retrospective MWAS to identify new prognostic biomarkers. Given the complementarity of MWAS and GWAS, future molecular epidemiological surveys should be able to capture both genetic and environmental information providing fresh insights into changing disease patterns around the world and informing personalized and public healthcare strategies.

Despite the potential for metabolic phenotyping, there remain many challenges both analytical and biological. The sheer scale of screening required to obtain useful patient stratification information or epidemiological insights is still formidable, and there is still no established analytical exploration strategy. However, we do know that a range of techniques both "shotgun" and targeted will be required to give the metabolome coverage needed to discover new biomarker combinations for disease risk factors in the general population. But these are mainly technical challenges and will inevitably fall with the march of analytical discovery. Perhaps the greatest challenge of all will be data visualization, that is, capturing the richness of mathematical models of the metabolic phenotypes of vast numbers of people so that their biological messages can be understood by the medical practitioners of the future.

## REFERENCES


Clayton, T.A., Lindon, J.C., Antti, H., Charuel, C., Hanton, G., Provost, J.P., Le Net, J.L., Baker, D., Walley, R.J., and Everett, J.E. (2006). Nature *440*, 1073–1077.

Chen, Y., Zhu, J., Lum, P.Y., Yang, X., Pinto, S., MacNeil, D.J., Zhang, C., Lamb, J., Edward, S., Sieberts, S.K., et al. (2008). Nature *452*, 429–435.

Daviglus, M.L., Stamler, J., Pirzada, A., Yan, L.L., Garside, D.B., Liu, K., Wang, R., Dyer, A.R., Lloyd-Jones, D.M., and Greenland, P. (2004). JAMA *292*, 1588–1592.

Dumas, M.E., Wilder, S.P., Bihoreau, M.T., Barton, R.H., Fearnside, J.F., Argoud, K., D'Amata, L., Wallis, R.H., Blancher, C., Keun, H.C., et al. (2007). Nat. Genet. *39*, 666–672.

Holmes, E., Loo, R.L., Stamler, J., Bictash, M., Yap, I.K., Chan, Q., Ebbels, T., De Iorio, M., Brown, I.J., Veselkov, K.A., et al. (2008). Nature *493*, 396–401.

Kruglyak, L. (2008). Nat. Rev. Genet. *9*, 3114–3118.

Li, H., Wu, Y., Loos, R.J., Hu, F.B., Liu, Y., Wang, J., Yu, Z., and Lin, X. (2008a). Diabetes *57*, 264–268.

Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., Zhou, H., Zhang, Y., Shen, J., Pan, X., Zhang, M., et al. (2008b). Proc. Natl. Acad. Sci. USA *105*, 2117–2122.

Loscalzo, J., Kohane, I., and Barabasi, A.L. (2007). Mol. Syst. Biol. *3*, 124.

Nebert, D.W., Nebert, J., and Vessell, E.S. (2003). Am. J. Pharmacogenomics *3*, 361–370.

Nicholson, J.K. (2006). Mol. Syst. Biol *2*, 52.

Nicholson, J.K., Connelly, J., Lindon, J.C., and Holmes, E. (2002). Nat. Rev. Drug Disc. *1*, 153–161.

Schadt, E.E., Lamb, J., Yang, X., Zhu, J., Edwards, S., Guhathakurta, D., Sieberts, S.K., Monks, S., Reitman, M., Zhang, C., et al. (2005). Nat. Genet. *37*, 710–717.

Sindic, A., Chang, M.H., Mount, D.B., and Romero, M.F. (2007). Curr. Opin. Nephrol. Hypertens. *16*, 484–490.

Westhof, M., Litterst, P., Freitag, L., and Baumbach, J.I. (2007). J. Physiol. Pharmacol. *58* (*Suppl 5*), 739–751.